Fall 2019

Math 605: Numerical Methods in Stochastic Analysis

The idea is to (1) do the problems, (2) think about the questions, (3) have a one-sentence description of the each of the key ideas and "other points", (4) remember some of the "random bits and pieces".

Problems

(1) Compare five ways to generate a standard normal random variable: (a) acceptance-rejection with exponential, (b) Box-Muller polar method, (c) Marsaglia polar method, (d) ratio of two uniforms, (e) the build-in method in MATLAB or similar computer algebra system. Compare both the **computational complexity** (e.g. using the CPU time or the number of flops) and the **quality** (using, e.g. Q-Q plot, χ^2 test, or some measure of the difference between the empirical and the normal cdf).

(2) Generate 100 iid symmetric α-stable random variables corresponding to $\alpha = 0.1$ and to $\alpha = 0.9$. In each case, run your favorite statistical test confirming (or denying?) that you got the right distribution.

(3) Implement a procedure for generating iid Poisson random variables, and a procedure for generating iid geometric random variables. In each case, make sure to include the corresponding parameter.

(4) Let $X_1, \ldots, X_n$ be iid copies of a random variable X and let $Y_1, \ldots, Y_n$ be iid copies of a random variable Y, independent of X. Assume that both X and Y have finite variance. Investigate the following two ways to estimate the product $(\mathbb{E}X) \cdot (\mathbb{E}Y)$: as a product of the sample means

$$\left(\frac{1}{n}\sum_{k=1}^{n} X_k\right) \cdot \left(\frac{1}{n}\sum_{k=1}^{n} Y_k\right)$$

and as the sample mean of the product

$$\frac{1}{n}\sum_{k=1}^{n} X_k Y_k.$$

Namely (a) determine analytically which estimator has smaller variance; (b) check your conclusion experimentally (for example, when X is normal with mean 1 and variance 1 and Y is uniform on $[0, 1]$; you are welcome to consider other distributions. The choice of n is up to you).

(5) Let $X = (X_0, X_1, X_2, \ldots)$ be a mean-square stationary sequence, that is, $\mathbb{E}X_n = \theta$ and $\mathrm{Cov}(X_{k+n}, X_k) = \mathrm{Cov}(X_n, X_0) = \rho(n)$ for all $n, k \geq 0$. For $r = 1, 2, 3, \ldots$, define

$$\hat{\theta}_{n,r} = \frac{1}{n}\sum_{k=1}^{n} X_{kr}, \hat{\theta}_{nr} = \frac{1}{nr}\sum_{k=1}^{nr} X_k.$$

Each of $\hat{\theta}_{n,r}$ and $\hat{\theta}_{nr}$ can be considered an estimator of θ.

(a) Compute the mean-square error for each estimator. How does the mean square error depend on r? Which estimator is the best?

(b) Find a sufficient condition for the estimators to be consistent as $n \to \infty$.

(c) Check your conclusions experimentally when $X_k = aX_{k-1} + \xi_k$, where $|a| < 1$, $\xi_k, k \geq 1$, are iid standard normal, and X_0 is normal with mean zero and variance $1/(1-a^2)$ [why?], independent of ξ_k. Try both positive and negative values of a, for example, $a = \pm 1/2$.

(6) Implement Siegmund's algorithm when the random walk is made up of (a) iid Bernoulli random variables with values ± 1 with the corresponding probabilities 0.4 and 0.6; (b) iid normal random variables with mean -0.2 and variance 0.24. In both cases, you need some preliminary analytical work to get the optimal value of the parameter for the change of measure, which involves computation of the moment generating function. To solve the corresponding equation, you are welcome to use computer. Try several different values for the level x. Also, see if you can get analytical bound on the corresponding probability of the first passage: this will give you an idea how large x should be in order for the change of measure to make computational sense.

(7) Check how Example 1.1 in Chapter VI of the book works on computer, taking $\pi = 0.00001$ and $m = 10$. More specifically, the problem is to estimate the probability that, in Bernoulli trials with probability of success in one trial 10^{-5}, the first success will happen during the first 10 trials. This probability is very small: the probability to have no successes in 10 trials is $(0.99999)^10 = 0.999900005$ (according to my pocket calculator), and so we are looking at the probability 0.000099995, which is pretty close to the approximation $m\pi = 10^{-4}$. Your task is to estimate this probability to within 5% relative accuracy with 95% confidence [that is, the total length of the 95% confidence interval for π should be no more than 10% of π, which is 10^{-5}] using Monte Carlo simulations of Bernoulli trials with probability of success in one trial 0.1. The exact formula to use is on top of page 161 of the book. Also, estimate how many simulations you expect to conduct and compare the result with (a) the actual number, (b) the estimated number of experiments in the case of direct simulation (when you expect about 10,000 experiments to get a single "success").

For extra credit, think about the following: since the procedure works the same way for any probability of success, why not take probability of success bigger (say, 0.5), so that the probability of the event in question is closer to 1? You are welcome to try the corresponding experiment with equally likely success and failure and confirm that the results are not as good. You are even more welcome to provide mathematical reasons for that.

(8) Investigate how the Robbins-Monro algorithm works for finding zeroes of the function $f(x) = \sin x$ observed in standard Gaussian noise.

(9) Determine all zeros and critical points of the function

$$f(x) = (x-2)(x-1)x^2(x+1)^3(x+2)^3$$

observed in standard Gaussian noise. What if the observation noise is symmetric Cauchy?

(10) Reproduce the pictures from Figures 3.1 and 3.2 on page 266 in the book. Of course, the exact reproduction is impossible, given the random nature of the experiment, so the objective is to evaluate the integral on page 265 using (a) standard Monte Carlo method; (b) Quasi-Monte Carlo method with a Halton

sequence, and to plot the approximations against the number of simulations. Beside the Halton sequence, you are welcome to try other low discrepancy sequences.

(11) Reproduce, as much as possible, the pictures from page 303 of the book, and also make a picture of the sample path of the process $W(t)-t$ with reflection at zero. For extra credit, confirm experimentally that the stationary distribution of the process is exponential with mean $1/2$.

(12) Generate a sample path of the Lévy process with the triplet $(0, 0, \nu)$, where $\nu(dx) \sim |x|^{-\alpha-1}dx$ (symmetric stable) or $\nu(dx) \sim x^{-\alpha-1}I(x > 0)\,dx$ (the corresponding subordinator) for $\alpha = 1/2, 1, 3/2$ (that would be six paths overall).

(13) Design a procedure generating a prescribed number of prescribed orthogonal polynomials. Test it on Hermite and something else.

(14) Design a procedure generating a prescribed number of prescribed Appel polynomials. Confirm that, under the right conditions, you get the same Hermite polynomials as in the orthogonal case.

Questions

(1) How to use nonlinear recursions (e.g. logistic map) to generate random numbers?

(2) What are the correct normalization and the limit distribution for the L_p norm of the difference between the empirical and true cdfs, for $1 \le p < \infty$? The Kolmogorov-Smirnov test corresponds to $p = \infty$.

(3) How to use the function $F^{\leftarrow}$ to generate discrete random variables (e.g. Poisson)?

(4) How to quantify the idea of the von Mises-Church collective?

(5) How to generate a uniform distribution on a manifold $G \subset \mathbb{R}^n$ when the Lebesgue measure of G is zero? The famous example (mentioned in the text) is the 2-dim sphere in the three-dimensional space. Can we extend it to an ellipsoid? Will it be easier to generate uniform distribution on the torus, the Klein bottle, or the real projective plane?

(6) In what models can we get convergence rate faster than the canonical $n^{1/2}$, and how much faster can this rate be?

(7) How to modify the jackknife procedure by omitting more than one observation at a time, and can this (or some other) modification lead to higher-order bias reduction?

(8) What is the general form of the expansion of the bias and what are the conditions for the expansion to hold up to a prescribed order?

(9) How does sectioning procedure work for the particular example of estimating the population mean and is it the same as using the idea of sectioning in the finite-dimensional (as opposed to infinite-dimensional or functional) setting?

(10) If a CLT holds for the sample mean of $f(X(t))$, where $X = X(t)$ is an ergodic process, then $\int_0^\infty \mathrm{Cov}_\pi\big(f(X(t)), f(X(0))\big)dt$ is finite and positive, where Cov_π means the covariance is computed under the assumption that X is stationary and $X(t)$ has the invariant distribution π for all $t \ge 0$. What conditions on the

process and the function f ensure that the integral is indeed finite and positive? Will those conditions be enough to imply the CLT? Are there examples of an ergodic process for which the integral is (a) finite but negative, (b) positive infinite (c) negative infinite, (d) does not exist at all? How does the function f enter the picture? The same questions apply to discrete time too, and could be easier to study.

(11) Derivation of the regenerative ratio formula used the solution of a certain equation, and so the underlying assumption is that the equation has at least one solution; this assumption is non-trivial if the state space of the process is infinite. While the solution of the equation does not appear in the final formula, what happens to the formula if the equation does not have a solution? Can it happen that the solution is such that the corresponding process is not a martingale but only a local martingale?

(12) Ideal importance sampling looks very much like size biasing, which suggests potential connections with Stein's method. Can you think of any useful connections?

(13) How small can the probability $\mathbb{P}^*(S_n > nx)$ be after the optimal exponential tilting of measure (e.g., do we have a uniform (in n and/or x) bound from below)?

(14) How will the Robbins-Monro algorithm work with perfect evaluations of the function?

(15) How will the secant method work with noisy measurements of the function?

(16) What other recursive algorithms can benefit from a Polyak-Ruppert-type averaging?

(17) Is bisection method equivalent to generating Brownian motion using anti-derivatives of the Haar basis functions? Will higher-order wavelets make corresponding simulations any better?

(18) Why do both strong and weak convergence orders consider the values only at the terminal time? How will the analysis and results change if a function-space norm is considered instead?

(19) The Lévy process generalizes the standard Brownian motion: similar to the Brownian motion, we get stationary and independent increments, but, unlike the Brownian motion, the trajectories are not necessarily continuous. What would be a similar generalization of the fractional Brownian motion? For some results in this direction, see *A Unifying Approach To Fractional Lévy Processes* by E. Engelke and J. H. C. Woerner in Stochastics and Dynamics Vol. 13, No. 2 (2013).

(20) Consider the circulant embedding of the covariance matrix of the increments of the fractional Brownian motion. Is the matrix always positive definite? The

cases to check are (a) different values of the Hurst parameter and (b) a possibility to have non-uniform spacing of the time points where we sample the process.

(21) What can we say about a *random* circulant embedding matrix? In other words, in the original circulant embedding of the correlation matrix, replace the distinct entries with iid random variables supported in $[-1, 1]$. Will the result be positive definite?

Key ideas

(1) Acceptance-rejection method.
(2) Copula (to model dependence and to extend distributions other than normal to several dimensions).
(3) Canonical rate of convergence and variance control.
(4) Confidence interval.
(5) Bias and mean-square error.
(6) Delta method.
(7) Invariant distribution.
(8) Regeneration.
(9) Perfect sampling.
(10) Relaxation time.
(11) Importance sampling.
(12) Antithetic sampling.
(13) Stratification.
(14) Bounded relative error vs. logarithmic efficiency.
(15) Rare event simulation via importance sampling.
(16) Siegmund's algorithm.
(17) Doob's h transform for Markov processes (a general procedure to condition a Markov process on something).
(18) Robbins-Monro and Kiefer-Wolfowitz algorithms.
(19) Using a low-discrepancy sequence to speed up Monte Carlo simulations.
(20) Weak and strong order of approximation for stochastic differential equations.
(21) Milstein's scheme.
(22) Simulation of a Lévy process.
(23) Metropolis-Hastings algorithm.
(24) Simulating a Gaussian vector using the circulant embedding of its covariance matrix and FFT.
(25) Iterative Function System (IFS) representation of a Markov chain.

Other points

(1) The left-continuous inverse of a cdf.
(2) Generation of a random variable as a ratio of two dependent uniforms.
(3) Glivenko-Cantelli theorem and related results about convergence of the empirical cdf.
(4) Bootstrapping.
(5) Jackknifing.
(6) Sectioning.
(7) Confidence ellipsoid.
(8) Simulation budget.
(9) CLT for the sample mean for dependent random variables.
(10) Propp-Wilson algorithm.
(11) Variance reduction by conditioning.
(12) Exponential family and exponential tilting of measure.
(13) Heavy tails vs. light tails.
(14) Large deviations bound vs. sharp asymptotic.
(15) Rare event simulation along the most likely of all the unlikely trajectories (the true Large Deviations approach to rare event simulation).
(16) How to estimate the tail of a distribution.
(17) Estimating the derivative (FD, IPA, and LR methods).
(18) Polyak-Ruppert averaging.
(19) EM algorithm.
(20) "Asymptotic" analysis of the Robbins-Monro algorithm in one dimension.
(21) Different ways to generate Brownian motion and the L^1 and L^∞ approximation errors for the Gaussian random walk/linear interpolation method.
(22) From a stochastic differential equation with additive noise to a random equation.
(23) Fractional Brownian motion.
(24) Special types of the Lévy processes (compound Poisson, symmetric stable, subordinator).
(25) Component-wise Metropolis-Hastings algorithm and the Gibbs sampler.
(26) Using the birth-death process to represent a number of basic queuing systems, such as M/M/m and M/M/∞.

Random bits and pieces

(1) Log-convex/concave functions.
(2) von Mises-Church collective.
(3) Poisson point process.
(4) Markov, semi-Markov, and generalized semi-Markov processes.
(5) Frechet-Hoeffding bound.
(6) Law of iterated log and sequential analysis.
(7) Buffon's needle.
(8) Perpetuity.
(9) Implied volatility.
(10) Skewness and kurtosis.
(11) Palm inversion.
(12) Harris chain.
(13) Latin square.
(14) Call-put parity.
(15) Variance decompositions.
(16) Some definitions and results from ergodic theory.
(17) Function of regular variation.
(18) Mogul'skii's theorem.
(19) Strong and weak laws of large numbers.
(20) Greeks (in option pricing).
(21) First-order linear finite-difference equation with variable coefficients.
(22) P vs NP problem.
(23) Parameter estimation with heavy-tailed noise.
(24) Quadrature rules of Newton-Cotes and Gauss.
(25) The Bessel process.
(26) The Langevin equation and the role of $\sqrt{2}$ in the diffusion.
(27) Circulant (matrix).
(28) Self-similarity.
(29) FFT.
(30) A 4-th order rational approximation of the inverse of the standard normal cdf.
(31) Domination number of a graph.
(32) Failures in time (FIT) as a measure of reliability.